(19) BUNDESREPUBLIK DEUTSCHLAND

DEUTSCHES PATENTAMT

(12) **Offenlegungsschrift**

(11) **DE 3929481 A1**

(51) Int. Cl. 5: G 10 L 3/00

(21) Aktenzeichen: P 39 29 481.1
(22) Anmeldetag: 5. 9. 89
(43) Offenlegungstag: 15. 3. 90

DE 3929481 A1

(30) Unionspriorität: (32) (33) (31)
07.09.88 JP 63-222309

(71) Anmelder:
Hitachi, Ltd., Tokio/Tokyo, JP

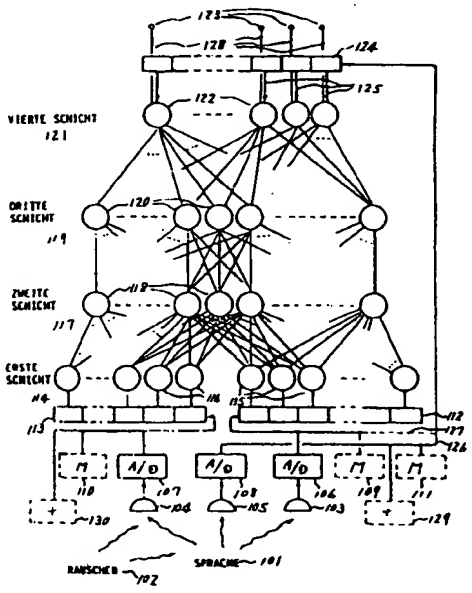
(74) Vertreter:
Strehl, P., Dipl.-Ing. Dipl.-Wirtsch.-Ing.; Schübel-Hopf, U., Dipl.-Chem. Dr.rer.nat.; Groening, H., Dipl.-Ing., Pat.-Anwälte; Schulz, R., Dipl.-Phys. Dr.rer.nat., Pat.- u. Rechtsanw., 8000 München

(72) Erfinder:
Ichikawa, Akira, Musashino, Tokio/Tokyo, JP;
Asakawa, Yoshiaki, Kawasaki, Kanagawa, JP;
Amano, Akio, Higashimurayama, Tokio/Tokyo, JP;
Hataoka, Nobuo, Pittsburgh, Pa., US

Prüfungsantrag gem. § 44 PatG ist gestellt

(54) Verfahren und Vorrichtung zur Vorbearbeitung von Sprachsignalen

Der zu dem Verfahren und der Vorrichtung zur Vorbearbeitung von Sprachsignalen verwendete Sprachdatenfilter enthält eine Anzahl von Mikrofonen (103, 104, 105), die im Abstand voneinander angeordnet sind. Der auf die Mikrofone einwirkende Schall wird in nachgeschalteten A/D-Konvertern (106, 107, 108) in ein digitales serielles Signal umgewandelt, das ein Eingangssignal für ein neuronales Netzwerk bildet. Das neuronale Netzwerk blendet Hintergrundgeräusche aus, wobei teilweise Daten verwendet werden, die aus den Parallax-Informationen erhalten werden, die durch die versetzte Anordnung der Mikrofone gewonnen werden. Die aus dem neuronalen Netzwerk erhaltenen Daten werden dann zu einem digitalen Signalprozessor übertragen, um das Rauschen herauszufiltern.



DE 3929481 A1

Beschreibung

Die Erfindung betrifft allgemein die Signalverarbeitung und insbesondere ein Verfahren und eine Vorrichtung zur Vorbearbeitung von Sprachsignalen, um den Rauschabstand bei den einem Sprachprozessor zugeführten Sprachsignalen zu verbessern.

Es sind einige Verfahren zur Verbesserung des Rauschabstandes bei Sprachsignalen bekannt, wobei die Frequenzeigenschaften des Rauschens vorab untersucht werden, um die Rauschkomponente dann vom Sprachsignal subtrahieren zu können. Diese bekannte Verfahren beruhen jedoch auf der falschen Annahme, daß das Hintergrundrauschen (das Hintergrundgeräusch) gleichmäßig ist. Solches Systeme arbeiten typisch mit zwei Mikrofoneingängen, um die entsprechenden Signale substrahieren zu können und dadurch das Hintergrundrauschen auszublenden. Es wurde auch bereits die Verwendung eines sogenannten neuronalen Netzwerkes diskutiert (Proceedings of ASJ (Acoustic Society of Japan) Spring Meeting, 3-p-13, Seiten 253 bis 294, Mai 1988).

Das aus der letztgenannten Druckschrift bekannte System zeigt eine verbesserte Leistungsfähigkeit und hat einen Rauschabstand, der dem früherer Techniken überlegen ist, es hat sich jedoch herausgestellt, daß dabei die Verständlichkeit herabgesetzt ist.

Die Bezeichnung "neuronales Netzwerk" schließt hier zwei Arten der neuronalen Netzwerk ein. Bei der ersten Art besteht das neuronale Netzwerk aus gleichwertigen parallel verarbeitenden Elementen, die untereinander entsprechend einer dynamisch selbstorganisierenden Programmierung auf eine nichtüberwachte, das heißt selbstlernende Weise verbunden werden, unabhängig davon, ob ein "Lehrer" vorhanden ist oder nicht. Bei der zweiten Art des neuronalen Netzwerkes besteht das Netzwerk aus gleichwertigen, parallel verarbeitenden Elementen, die vorab durch Lernen fest miteinander verbunden werden. Ein solches Netzwerk kann dann später nichts mehr "lernen".

Die menschliche Sprache wird aus dem Mund als eine Folge von Verdichtungen und Verdünnungen der Luftmoleküle abgegeben. Die sprachbildenden Organe, über die die Sprachinformationen ausgegeben werden, sind bei jedem Menschen anders. Durch die physikalischen Unterschiede zwischen den einzelnen Menschen weichen die physikalischen Eigenschaften der Sprachsignale, wenn sie als physikalische Signale betrachtet werden, erheblich voneinander ab. Darüber hinaus wird von den verschiedensten Schallquellen aus den verschiedensten Richtungen Rauschen oder ein Hintergrundgeräusch erzeugt. Die Abweichungen in den physikalischen Eigenschaften von Sprachsignalen weisen daher keine Gemeinsamkeiten auf.

Aufgabe der Erfindung ist es, ein Verfahren und eine Vorrichtung zu schaffen, mit dem bzw. mit der Sprachdaten mit verbesserter Verständlichkeit und Klarheit erhalten werden können.

Zur Lösung dieser Aufgabe werden bei dem erfindungsgemäßen Verfahren bzw. der entsprechenden Vorrichtung die Signale einer Anzahl von Wandlern, die Schall in elektrische Signale umwandeln, wie beispielsweise Mikrofone, als Eingangssignale eines neuronalen Netzwerkes verwendet. Die sprachliche Konversation wird vom Menschen ja ohne Schwierigkeiten ausgeführt, auch bei einem hohen Geräuschpegel, teilweise durch die Benutzung von beiden Ohren. Die Anzahl von Mikrofonen ergibt Eingangsinformationen, wie etwa Parallax-Informationen, die das neuronale Netzwerk verwenden kann, um eine Schallfilterung auszuführen.

Das neuronale Netzwerk führt einen Lernvorgang derart aus, daß nur physikalische Eigenschaften, die den Eingangssignalen von einer Anzahl von Mikrofonen und einem reinen Sprachsignal, das zum Lernen von der Ausgangsseite des Netzwerkes zugeführt wird, gemeinsam sind, durchgelassen werden. Alle anderen Signale werden ausgefiltert. Es werden somit nur Signale durchgelassen, die ausschließlich die physikalischen Eigenschaften der Sprache aufweisen, während die Rauschkomponente unterdrückt wird. Der Rauschabstand des Systems wird dadurch wesentlich verbessert.

Mit dem erfindungsgemäßen Verfahren und der entsprechenden Vorrichtung ist es möglich, den Rauschabstand von Sprachinformationen zu erhöhen, die bei einem sehr hohen Geräuschpegel erfaßt werden. Es werden somit die folgenden Vorteile erhalten:

Die Sicherheit der Spracherkennung wird dadurch erhöht, daß ein erfindungsgemäßer Filter vor einer Spracherkennungsvorrichtung angeordnet wird.

Die Sicherheit der Erkennung wird auch dadurch erhöht, daß der erfindungsgemäße Filter vor einer Sprachkodiervorrichtung angeordnet wird, wodurch eine kodierte Sprache mit einem hohen Rauschabstand, die leicht zu erkennen ist, erhalten wird, so daß eine Sprachkommunikation auch bei hohem Geräuschpegel ausgeführt werden kann.

Die Sicherheit der Erkennung wird auch dadurch erhöht, daß der erfindungsgemäße Filter vor einem der gewöhnlichen verschiedenen Arten von Sprachanalysegeräten angeordnet wird, wodurch es möglich ist, Verzerrungen der Sprache bei einem hohen Geräuschpegel festzustellen (unter solchen Bedingungen wird von einem Menschen die Stimme im allgemeinen angehoben, um das Rauschen zu übertönen, wodurch die Sprache von ihrer gewöhnlichen Form abweicht).

Mit dem erfindungsgemäßen System ist es daher möglich, den Rauschabstand bei Sprachdaten zu erhöhen, ohne daß gleichzeitig die Verständlichkeit verschlechtert wird.

Ein Ausführungsbeispiel des erfindungsgemäßen Systems wird im folgenden anhand der Zeichnung näher erläutert. Es zeigt

Fig. 1 den Aufbau eines Filters zur Verbesserung des Rauschabstandes mit einer Darstellung des Lernprozesses, der damit verbunden ist;

Fig. 2 eine Vorrichtung, bei der der Filter der **Fig.** 1 verwendet wird; und

Fig. 3 Beispiele zur Verwendung des erfindungsgemäßen Systems.

In der **Fig.** 1 ist die Anordnung eines Filters zur Verbesserung des Rauschabstandes, der ein neuronales Netzwerk enthält, dargestellt. Die **Fig.** 2 zeigt die Anwendung dieses Filters bei einem filternden und lernenden System.

Das in der **Fig.** 1 gezeigte neuronale Netzwerk besteht aus einer Anzahl von "Neuronen", die in einer ersten bis vierten Schicht **114, 117, 119** und **121** angeordnet sind. Wie es allgemein bekannt ist, können die einzelnen Neuronen durch Verarbeitungseinheiten gebildet werden, die eine Bewertung oder Gewichtung der Signale an ihrem Eingang vornehmen, oder sie können durch eine herkömmliche Von-Neuman-Maschine emuliert werden. Zum Aufbau des Netzwerkes können selbstverständlich auch mehr oder weniger Neuronen und/oder Schichten bzw. Ebenen wie in der **Fig.** 1 verwendet werden.

Wie in der Fig. 2 gezeigt, werden Sprachsignale 101 und Rauschsignale 102, die einer Anzahl von Mikrofonen 201 eingegeben werden, durch einen multiplexenden A/D-Konverter 202 digitalisiert und dann zu einem Schalter 203 geführt. Im Lernmodus wird das digitalisierte Signal vom Schalter 203 zu einem internen Bus 204 geführt und unter der Steuerung eines Mikroprozesses (μ-CPU) 205 in einem Speicher 206 gespeichert, um entsprechend der im Mikroprozessor 205 enthaltenen Prozeduren das neuronale Netz aufzubauen. Das Ergebnis des Lernens wird in der Form von Gewichtungsfaktoren für die Verbindungen zwischen den Elementen des neuronalen Netzes erhalten.

Jeder so bestimmte Gewichtungsfaktor wird über einen Signalleitung 207 zu einem digitalen Signalprozessor (DSP) 208 gegeben, der ein neuronales Netzwerk zur Rauschfilterung enthält. Der digitale Signalprozessor 208 stellt somit ein neuronales Netzwerk zur Rauschfilterung dar, bei dem die Gewichtungen bereits festgelegt ("gelernt") sind. Wenn das System als Rauschfilter verwendet wird, wird das Eingangs-Sprachsignal 101 (und das Rauschsignal 102) direkt über die Mikrofone 201, den A/D-Konverter 202 und den Schalter 203 in den Signalprozessor 208 eingegeben, um ein Signal 209 mit verbessertem Rauschabstand am Ausgang des Prozesses 208 zu erhalten. Wenn die Anordnung nur als Rauschfilter verwendet wird, brauchen die zum Lernen benötigten Elemente natürlich nicht immer vorhanden zu sein.

Die Arbeitsweise dieses Rauschfilters und die Lernprozedur wird anhand der Fig. 1 beschrieben. Beim Lernen können einige der in der Fig. 1 gezeigten Teile durch virtuelle Teile des Mikroprozessors 205 und des Speichers 206 der Fig. 2 realisiert werden, während sich bei der Ausführung einer tatsächlichen Operation nur diejenigen Teile im Signalprozessor 208 befinden, die den in den Fig. 1 gezeigten Filter bilden. Es ist natürlich auch möglich, daß sich die Mikrofone 201 und der A/D-Konverter 202 an einem anderen Ort befinden und über eine digitale Leitung mit dem Signalprozessor 208 verbunden sind, der dann allein die Vorrichtung bildet.

Zur Vereinfachung erfolgt die Beschreibung mit Bezug auf eine Anordnung, die zwei Eingangssysteme beinhaltet. Die Anordnung kann jedoch auf die gleiche Weise auch drei oder mehr Eingangssystem beinhalten.

Bei der Darstellung der Fig. 1 ist angenommen, daß das Ausgangssignal des q-ten Neuron-Elementes in der p-ten Schicht gleich $O_{p,q}$ und das Ausgangssignal des r-ten Elementes in der $(p-1)$-ten Schicht gleich $O_{p-1,r}$ ist. Zur Vereinfachung der Beschreibung wird weiter angenommen, daß die Übertragungseigenschaft zwischen dem Eingang x und dem Ausgang y für alle Elemente gleich ist und dargestellt wird durch

$$y = f(x). \quad (1)$$

Dann gilt folgendes:

$$I_{p,q} = \Sigma w_{p-1,q,r}(O)_{p-1,r} \quad (2)$$

$$O_{p,q} = f(I_{p,q}). \quad (3)$$

Aus der Gleichung (2) ist ersichtlich, daß die Verarbeitung viele Berechnungen zur Bildung der Summe von Produkten beinhaltet, die der Signalprozessor ausführt. Das neuronale Netzwerk beinhaltet vorzugsweise eine große Anzahl von Neuron-Elementen, die die durch die Gleichung (3) ausgedrückte Eigenschaften haben und die miteinander in einer hierarchischen Struktur verbunden sind. Es ist anzumerken, daß, obwohl das in der Fig. 1 gezeigte neuronale Netzwerk aus vier Schichten oder Ebenen besteht, die Anzahl der Schichten nicht notwendigerweise auf vier begrenzt ist.

Das Gemisch aus den Sprachsignalen 101 und dem Rauschen 102, das über die Mikrofone 103 und 104 den A/D-Konvertern 106 und 107 zugeführt wird, wird dort in digitale Signale umgewandelt, die zu Schieberegistern 112 bzw. 113 weitergeleitet werden. Die Schieberegister 112 und 113 sind zusammen mit einem Schieberegister 124 (später noch genauer erläutert) dafür vorgesehen, aufeinanderfolgend die Daten synchron zur Abtastperiode der A/D-Konverter zu verschieben und in jeder Stufe Daten auszugeben. Die Ausgangssignale der verschiedenen Stufen der Schieberegister 112 und 113 werden dann jeweils den Elementen 115 bzw. 116 in der ersten (Eingangs-) Schicht 114 des neuronalen Netzwerkes zugeführt.

Die Ausgangssignale der Elemente 115 und 116 der ersten Schicht werden auf der Basis der durch die Gleichungen (2) und (3) ausgedrückten Beziehungen zu den Elementen 118 der zweiten Schicht 117 weitergeleitet. Das gleiche gilt für die Verbindung zwischen den Elementen 118 in der zweiten Schicht 117 und den Elementen 120 in der dritten Schicht 119 sowie der Verbindung zwischen den Elementen 120 in der dritten Schicht 119 und den Elementen 122 in der vierten (Ausgangs-) Schicht 121. Durch die Verarbeitung der Signale in den Elementen auf der Basis der Beziehungen, die durch die Gleichungen (2) und (3) dargestellt werden, werden Signale 128 mit einem verbesserten Rauschabstand an den Ausgangsanschlüssen 123 der Ausgangsschicht 121 abgegeben. Wenn das Ausgangssignal von einem der Ausgangsanschlüsse 123 als externes Ausgangssignal herausgenommen wird, wird ein Ausgangs-Sprachsignal 209 (Fig. 2) mit verbessertem Rauschabstand erhalten.

Es folgt eine Beschreibung des Lernvorganges bei dem neuronalen Netzwerk, das den Rauschfilter bildet.

Das rückwärtsschreitende Verfahren, das bei der Architektur neuronaler Netzwerke bekannt ist, wird für den Lernvorgang bei dem vorliegenden System geeignet angewendet. Ein solches rückwärtsschreitendes Verfahren ist beispielsweise in der Literaturstelle M.I.T. Press, "Parallel Distributed Processing" Band 1 (1986), Kap. 8, Seiten 318 bis 362 beschrieben.

Der Lernvorgang wird nun mit Bezug auf die Fig. 1 erläutert. Zur Vereinfachung werden einige Symbole eingeführt. Der Wert des Ausgangssignales 128 eines jeden Elementes 122 in der Ausgangsschicht 121 wird mit $O_{4,i}$ bezeichnet, der Wert des Ausgangssignales des j-ten Elementes der dritten Schicht 119 mit $O_{3,j}$, der Wert des Ausgangssignales des k-ten Elementes in der zweiten Schicht 117 mit $O_{2,k}$, und ein Soll-Ausgangswert, der als Lerneingang an das i-te Element in der vierten Schicht 121 angelegt wird, mit $T_{4,i}$. Bezüglich des Fehlersignales, das für jedes Signal im Verlauf des Rückwärtsschreitens erhalten wird, wird der Wert des Fehlersignals für das i-te Element in der vierten Schicht 121 mit $\delta_{4,i}$ der Wert des Fehlersignales für das j-te Element in der dritten Schicht 119 mit $\delta_{3,j}$, und der Wert des Fehlersignales für das k-te Element in der zweiten Schicht 117 mit $\delta_{2,k}$ bezeichnt. Des weiteren wird angenommen, daß die Übertragungseigenschaften der Elemente in allen Schichten die gleichen sind und derjenigen entsprechen, die durch die Gleichung (3) ausgedrückt wird. Es sei außerdem f' die Ableitung der Funktion f. Der Verbindungsfaktor zwischen dem i-ten Ele-

ment in der Ausgangsschicht 121 und dem j-ten Element in der dritten Schicht 119 wird mit $w_{3,i,j}$ bezeichnet, und der Verbindungsfaktor zwischem dem j-ten Element in der dritten Schicht 119 und dem k-ten Element in der zweiten Schicht 117 mit $w_{2,j,k}$.

Zum Lernen werden verschiedene Sprachtypen 101 und verschiedene Rauschtypen 102 getrennt in die Mikrofone 103, 104 und 105 eingegeben. Das in das Mikrofon 105 eingegebene Signal besteht aus einem reinen Sprachsignal, es wird für den Soll-Ausgangswert $T_{4,i}$ verwendet. Die Signale werden in den jeweiligen Speichern 109, 110 und 111 (Bereiche im Speicher 206 der Fig. 2) gespeichert. Die gespeicherte Sprache und das gespeicherte Rauschen werden in Addierern 129 und 130 addiert, um Signale zusammenzusetzen, denen ein Rauschen überlagert ist. Diese Signale werden zu den Schieberegistern 112 und 113 gegeben. Daten über das Ausmaß, in dem das Rauschen überlagert ist, und über die Kombination von Sprache und Rauschen werden für verschiedene erwartete Zustände wiederholt vorbereitet und als Lern-Eingangssignale verwendet. Bei der tatsächlichen Ausführung wird die Überlagerung durch Verwendung einer arithmetischen Funktion des Mikroprozessors 205 der Fig. 2 ausgeführt. Der Soll-Ausgangswert $T_{4,i}$ ist ein Sprachsignal, das dem Lern-Eingangssignal entspricht, und es setzt den Grad fest, bis zu dem die Sprache im Lern-Eingangssignal als Ergebnis der Verbesserung des Rauschabstandes verbessert werden soll. Das Eingangsmikrofon 105, der A/D-Konverter 108 und der Speicher 111 für den Soll-Ausgangswert $T_{4,i}$ können auch für einen Eingabe verwendet werden, das heißt als Mikrofon 103 (oder 104), D/A-Konverter 106 (oder 107) und Speicher 109 (oder 110), wie es durch die Verbindungslinie 127 gezeigt wird. Die Sprache für den Soll-Ausgangswert $T_{4,i}$ wird dem Schieberegister 124 eingegeben und die Ausgangssignale 125 aus den verschiedenen Stufen des Schieberegisters 124 werden den entsprechenden Elementen 122 in der Ausgangsschicht 121 des neuronalen Netzwerkes als Soll-Ausgangssignale 125 eingegeben.

Wenn jedem Element in der ersten Schicht 114 ein Lern-Eingangssignal (Sprache und Rauschen einander überlagert) eingegeben wird, wird auf der Basis der Beziehungen, die durch die Gleichungen (2) und (3) ausgedrückt werden, aufeinanderfolgend von jedem Element von der Eingangsschicht zur Ausgangsschicht ein Ausgangssignal erhalten. Nachdem das Ausgangssignal für jedes Element erhalten wurde, werden aufeinanderfolgend von der Ausgangsschicht 121 bis zu den unteren Schichten Fehlersignale ermittelt. Die Korrektur der Verbindungsfaktoren zwischen der p-ten Schicht und $(p+1)$-ten Schicht erfolgt unter Verwendung der Fehlersignale für die $(p+1)$-te Schicht und den Werten der Ausgangssignale in der p-ten Schicht. Im folgenden wird zur Vereinfachung nur der Vorgang zur Korrektur der Verbindungsfaktoren $w_{3,i,j}$ und $w_{2,j,k}$ erläutert. Für die folgenden Schichten wird der entsprechende Vorgang wiederholt, bis hinunter zur Eingangsschicht.

Zur Korrektur der Verbindungsfaktoren $w_{3,i,j}$ und $w_{2,j,k}$ werden den Wert $O_{2,k}$ des Ausgangssignales des k-ten Elementes in der zweiten Schicht 117, der Wert $O_{3,j}$ des Ausgangssignales des j-ten Elementes in der dritten Schicht 119, der Wert $\delta_{3,i,j}$ des Fehlersignales des j-ten Elementes in der dritten Schicht 119 und der Wert $\delta_{4,i}$ des Fehlersignales des i-ten Elementes in der vierten (Ausgangs-) Schicht 121 benötigt. Die Werte für $O_{2,k}$ und $O_{3,j}$ können durch eine Vorwärtsrechnung durch Anlegen von Eingangssignalen an die erste Schicht 114 erhalten werden, wie oben beschrieben. Die Werte $\delta_{4,i}$ $\delta_{3,j}$ können aus den folgenden Gleichungen berechnet werden:

$$\delta_{4,i} = (T_{4,i} - O_{4,i})\, f'\,(\sum_j w_{4,i,j}(O_{3,j})). \quad (4)$$

$$\delta_{3,j} = f'\,(\sum_k w_{2,j,k}(O_{2,k}))(\delta_{4,i})(w_{3,i,j}). \quad (5)$$

Als nächstes werden $w_{3,i,j}$ und $w_{2,j,k}$ korrigiert. Wenn die Korrekturwerte dabei durch $\Delta w_{3,i,j}$ und $\Delta w_{2,j,k}$ ausgedrückt werden, können diese Korrekturwerte wie folgt berechnet werden:

$$\Delta w_{3,i,j} = \alpha(\delta_{4,i})(O_{3,j}) \quad (6)$$

$$\Delta w_{2,k,h} = \alpha(\delta_{3,i})(O_{2,k}) \quad (7)$$

α kann durch experimentelles Überprüfen der Konvergenzgeschwindigkeit eingestellt werden. Die Gleichungen (6) und (7) ermöglichen eine Korrektur aller Verbindungsfaktoren zwischen der Ausgangsschicht und der dritten Schicht und zwischen der dritten und der zweiten Schicht. Die Verbindungsfaktoren zwischen der zweiten Schicht und der Eingangsschicht können auf die gleiche Weise korrigiert werden wie die Verbindungsfaktoren zwischen der dritten und der zweiten Schicht.

Auf diese Weise werden alle Verbindungsfaktoren einmal korrigiert. Mit anderen Eingangsdaten und Sollwerten (solchen, die sich von den obigen Werten bezüglich der Stimme, dem Rauschen, dem gegenseitigen Pegel und der gegenseitigen Phasenbeziehung unterscheiden) wird der obige Vorgang zur Korrektur der Verbindungsfaktoren wiederholt. Jedesmal, wenn der Vorgang wiederholt wird, wird ein Bewertungsfaktor E wie folgt ermittelt:

$$E = 1/2 \sum_i (T_{4,i} - O_{4,i})2. \quad (8)$$

Die Bewertungsfaktoren werden über alle Lernmuster gemittelt. Wenn der Mittelwert kleiner als ein vorgegebener Schwellenwert wird, steht fest, daß der Lernvorgang abgeschlossen ist.

Wenn der Standort einer sprechenden Person und die Positionen der Mikrophone auf einen vorgegebenen Bereich beschränkt sind, werden die Sprachinformationen zum Lernen auch unter den entsprechenden Bedingungen eingegeben und das interne Sprachsignal unter Berücksichtigung der Pegel und der Phasenunterschiede zwischen den Mikrofonen bei dieser Anordnung erzeugt. Dadurch wird die Effektivität der Verbesserung des Rauschabstandes bedeutend erhöht. Wenn für den Standort der sprechenden Person ein gewisser Bereich erlaubt werden soll, entspricht die Lern-Eingangssprache ebenfalls diesem Bereich. Die entsprechenden Bedingungen können leicht abgeleitet werden, beispielsweise auch durch eine interne Synthetisierung auf der Basis der Grundlagen der Akustik (beispielsweise reicht es, die Verzögerung des Sprachsignales, die sich aus dem Abstand zwischen der sprechenden Person und dem Mikrofon ergibt, und das quadratische Gesetz der Abschwächung zu berücksichtigen).

Es ist anzumerken, daß es auch möglich ist, das Eingangssignal einer komplexen Fourier-Transformation oder dergleichen zu unterwerfen und es dann beispiels-

weise im Frequenzraum in das neuronale Netz einzugeben. In einem solchen Fall kann die Eingangsschicht für die Frequenz und die Phase oder für den Realteil und den Imaginärteil in zweidimensionaler Form vorgesehen sein. Der Ausgang kann ein Ausgangssignal im Frequenzbereich sein, das in den Wellenformbereich rücktransformiert wird. Bei diesen Verfahren wird eine der bekannten Raumprojektionstransformationen und eine entsprechende inverse Transformation benötigt.

Einige Anwendungen des vorstehend beschriebenen Filters sind in der Fig. 3 dargestellt.

Die Spracherkennung kann beispielsweise dadurch verbessert werden, daß ein gemäß der vorstehenden Beschreibung aufgebauter Rauschfilter 301 vor einer Spracherkennungsvorrichtung 302 angeordnet wird, um daraus ein verbessertes Ausgangssignal 303 zu erhalten.

Der Rauschfilter 301 kann auch einer Sprachkodiervorrichtung 304 vorgeschaltet werden, wodurch an dessen Ausgang 305 eine kodierte Sprache erhalten wird, die leicht zu erkennen ist, so daß auch bei einem sehr hohen Geräuschpegel eine Sprachverbindung möglich ist.

Der Rauschfilter 301 kann schließlich auch vor einem gewöhnlichen Sprachanalysegerät 306 angeordnet werden, wodurch es möglich ist, Verzerrungen der Sprache bei einem hohen Geräuschpegel festzustellen, wenn beispielsweise von einem Menschen die Stimme angehoben wird, um die Hintergrundgeräusche zu übertönen, wodurch die Sprache von ihrer üblichen Form abweicht.

Der bei dem erfindungsgemäßen Verfahren bzw. der Vorrichtung zur Vorbearbeitung von Sprachsignalen verwendete Sprachdatenfilter weist somit eine Anzahl von Mikrofonen auf, die im Abstand voneinander angeordnet sind. Der auf die Mikrofone einwirkende Schall wird in nachgeschalteten A/D-Konverten in ein digitales serielles Signal umgewandelt, das ein Eingangssignal für ein neuronales Netzwerk bildet. Das neuronale Netzwerk blendet Hintergrundgeräusche aus, wobei teilweise Daten verwendet werden, die aus den Parallelax-Informationen erhalten werden, die durch die versetzte Anordnung der Mikrofone gewonnen werden. Die aus dem neuronalen Netzwerk erhaltenen Daten werden dann zu einem digitalen Signalprozessor übertragen, um das Rauschen herauszufiltern.

Patentansprüche

1. Vorrichtung zur Verringerung des Rauschens in Spracherkennungssystemen, **gekennzeichnet durch**

— eine Anzahl von räumlich getrennt angeordneten Wandlern (103, 104, 105; 201) zum Erzeugen einer Anzahl von elektrischen Sprachsignalen, die dem auf die Wandler einwirkenden Schall entsprechen;

— ein neuronales Netzwerk mit einer Anzahl von Schichten (114, 117, 119, 121), wobei jede Schicht aus einer Anzahl von Neuron-Elementen (115, 116; 118; 120; 122) besteht und die Schichten eine Eingangsschicht (114) und eine Ausgangsschicht (121) beinhalten;

— eine erste Kommunikationseinrichtung (106, 112), um das elektrische Sprachsignal von einem ersten (103) der Wandler zu jedem Element (115) eines ersten Satzes von Neuron-Elementen in der Eingangsschicht zu übertragen; und durch

— eine zweite Kommunikationseinrichtung (107, 113), um das elektrische Sprachsignal von einem zweiten (104) der Wandler zu jedem Element (116) eines zweiten Satzes von Neuron-Elementen in der Eingangsschicht zu übertragen.

2. Vorrichtung nach Anspruch 1, dadurch gekennzeichnet, daß jeder der Wandler (103, 104) eine Einrichtung zur Erzeugung eines analogen elektrischen Sprachsignales aufweist, das dem darauf einwirkenden Schall entspricht, und daß die erste und die zweite Kommunikationseinrichtung jeweils eine Einrichtung (106; 107) zur Umwandlung des analogen elektrischen Sprachsignales in ein erstes bzw. zweites serielles digitales Sprachsignal enthält.

3. Vorrichtung nach Anspruch 2, dadurch gekennzeichnet, daß die erste und die zweite Kommunikationseinrichtung jeweils ein Schieberegister (112; 113) zum Umwandeln der ersten bzw. zweiten seriellen digitalen Signale in eine entsprechende erste bzw. zweite Serie von Ausgangssignalen aufweisen, wobei die Ausgangssignale der ersten und zweiten Serien jeweils das Eingangssignal für ein Neuron-Element (115; 116) des ersten und zweiten Satzes der Eingangsschicht (114) bilden.

4. Vorrichtung nach Anspruch 3, gekennzeichnet durch einen Schalter (203) zum selektiven Anlegen eines reinen elektrischen Sprachsignales und eines gemischten elektrischen Sprach/Rauschsignales an das neuronale Netzwerk, und durch eine Einrichtung zum Ausführen eines überwachten Lernvorganges im neuronalen Netzwerk in Übereinstimmung mit dem zugeschalteten reinen elektrischen Sprachsignal und dem zusammengesetzten elektrischen Sprach/Rauschsignal, wodurch Neuron-Gewichtungsdaten erhalten werden, die die Übertragungseigenschaften zwischen den Neuron-Elementen des Neuronalen Netzes darstellen.

5. Vorrichtung nach Anspruch 3, gekennzeichnet durch eine Einrichtung (207) zum Übertragen der neuronalen Gewichtungsdaten vom neuronalen Netzwerk zu einem digitalen Signalprozessor, und durch einen digitalen Signalprozessor (208) zum Verarbeiten zusammengesetzter Sprach/Rauschsignale in Übereinstimmung mit den neuronalen Gewichtungsdaten.

6. Vorrichtung nach Anspruch 4, gekennzeichnet durch eine Einrichtung zur Ausführung einer Fourier-Transformation an wenigstens einem der elektrischen Sprachsignale, bevor dieses an das neuronale Netzwerk gegeben wird.

7. Vorrichtung nach Anspruch 4, gekennzeichnet durch

— eine Anzahl N zusätzlicher räumlich getrennter Wandler zur Erzeugung einer Anzahl von elektrischen Sprachsignalen, die einem darauf einwirkenden Schall entsprechen, wobei N eine positive ganze Zahl größer als Null ist;

— N zusätzliche Kommunikationseinrichtungen zum Übertragen des elektrischen Sprachsignales von jedem der N zusätzlichen Wandler zu jedem Element eines N-ten Satzes von Neuron-Elementen in der Eingangsschicht;

— wobei jeder der zusätzlichen N Wandler eine Einrichtung zum Erzeugen eines analogen elektrischen Sprachsignales entsprechend dem darauf einwirkenden Schall enthält;

— wobei die zusätzlichen N Kommunikations-

einrichtungen eine Einrichtung zur Umwandlung des analogen elektrischen Sprachsignales in das jeweilige N-te serielle digitalisierte Sprachsignal enthalten; und

— wobei jede der zusätzlichen N Kommunikationseinrichtungen Schieberegister zur Umwandlung der ersten bzw. zweiten seriellen digitalisierten Signale in die entsprechende N-te Serie von Ausgangssignalen enthält, wobei jedes Ausgangssignal der N-ten Serie ein Eingangssignal für ein Neuron-Element des N-ten Satzes der Eingangsschicht darstellt.

8. Verfahren zur Rauschverringerung in akustischen Signalen, gekennzeichnet durch die Verfahrensschritte

(a) des Aufnehmens von Schallwellen aus einer Anzahl von Positionen;

(b) des Erzeugens einer Anzahl von elektrischen Schallsignalen, die den Schallwellen von jeder der Positionen entsprechen;

(c) des Übertragens der elektrischen Schallsignale jeweils zu einem Satz von Neuron-Elementen (115; 116) in einer Eingangsschicht (114) eines neuronalen Netzwerkes,

(d) des Berechnens eines Ausgangssignales in dem neuronalen Netzwerk, das von den elektrischen Schallsignalen am ersten und zweiten Satz von Neutronen abgeleitet wird; und

(e) des Abgebens der Ausgangssignale von einer Ausgangsschicht (121) der Neuronen des neuronalen Netzwerkes.

9. Verfahren nach Anspruch 8, dadurch gekennzeichnet, daß der Verfahrensschritt (b) das Erzeugen einer Anzahl analoger elektrischer Sprachsignale und das Umwandeln der Anzahl analoger elektrischer Sprachsignale in eine entsprechende Anzahl serieller digitalisierter Sprachsignale umfaßt; und daß der Verfahrensschritt (c) die Übermittlung jedes der digitalen Sprachsignale zu dem entsprechenden Satz von Neuron-Elementen (115; 116) der Eingangsschicht (114) des neuronalen Netzwerkes beinhaltet.

10. Verfahren nach Anspruch 10, gekennzeichnet durch die weiteren Verfahrensschritte des selektiven Anlegens eines reinen elektrischen Sprachsignales und eines gemischten elektrischen Sprach/Rauschsignales an das neuronale Netzwerk und des Ausführens eines überwachten Lernvorganges im neuronalen Netzwerk in Übereinstimmung mit dem zugeschalteten reinen elektrischen Sprachsignal und dem zusammengesetzten elektrischen Sprach/Rauschsignal, wodurch Neuron-Gewichtungsdaten erhalten werden, die die Übertragungseigenschaften zwischen den Neuron-Elementen des neuronalen Netzes darstellen.

11. Verfahren nach Anspruch 10, gekennzeichnet durch die weiteren Verfahrensschritte des Übertragens der neuronalen Gewichtungsdaten von dem neuronalen Netzwerk zu einem digitalen Signalprozessor (208), des Übergebens gemischter Sprach/Rauschsignale vom neuronalen Netzwerk zum Signalprozessor; und des Verarbeitens der zusammengesetzten Sprach/Rauschsignale im Signalprozessor in Übereinstimmung mit den neuronalen Gewichtungsdaten.

12. Verfahren nach Anspruch 10, gekennzeichnet durch den weiteren Verfahrensschritte des Ausführens einer Fourier-Transformation an wenigstens einem der elektrischen Sprachsignale, bevor dieses an das neuronale Netzwerk gegeben wird.

13. Vorrichtung zur Verringerung von Signalrauschen, gekennzeichnet durch

— eine Anzahl räumlich getrennter Wandler (103, 104, 105; 201) zur Erzeugung einer Anzahl von elektrischen Sprachsignalen, die einem darauf einwirkenden Schall entsprechen;

— eine Anzahl von digitalisierenden Einrichtungen (106, 107, 108; 202) zum Umwandeln von analogen gemischten Sprach/Rauschsignalen in digitale Sprach/Rauschsignale;

— eine Einrichtung (207) zum Übertragen digital kodierter neuronaler Gewichtungsdaten von einem neuronalen Netzwerk zu einem digitalen Signalprozessor;

— einem digitalen Signalprozessor (208) zum Verarbeiten der digitalen gemischten Sprach/Rauschsignale von den Wandlern in Übereinstimmung mit den Gewichtungsdaten in gefilterte digitale Schalldaten; und durch

— eine Einrichtung zum Übertragen der gefilterten digitalen Schalldaten zu einer entsprechenden Empfangseinrichtung für die digitalen Schalldaten.

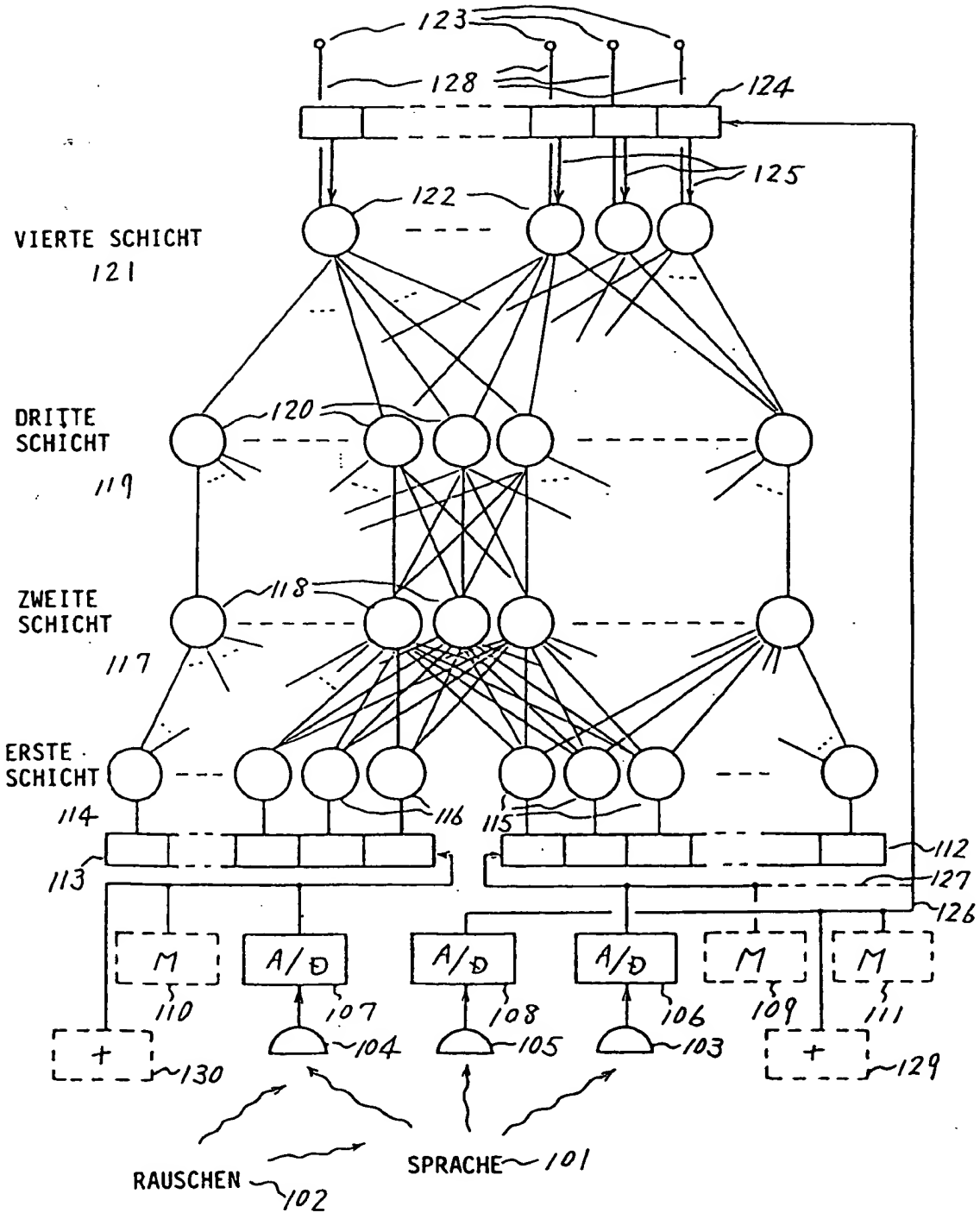
14. Vorrichtung nach Anspruch 13, dadurch gekennzeichnet, daß die Empfangseinrichtung für die digitalen Schalldaten eine Einrichtung zum Erzeugen analoger gefilterter Schalldaten aus den digitalen gefilterten Schalldaten aufweist.

15. Vorrichtung nach Anspruch 14, gekennzeichnet durch einen Lautsprecher zum Erzeugen gefilteter Schallwellen aus den analogen gefilterten Schalldaten.

Hierzu 2 Seite(n) Zeichnungen

Nummer: DE 39 29 481 A1
Int. Cl.[5]: G 10 L 3/00
Offenlegungstag: 15. März 1990

Fig. 1



Nummer: DE 39 29 481 A1
Int. Cl.[5]: G10L 3/00
Offenlegungstag: 15. März 1990

Fig. 2

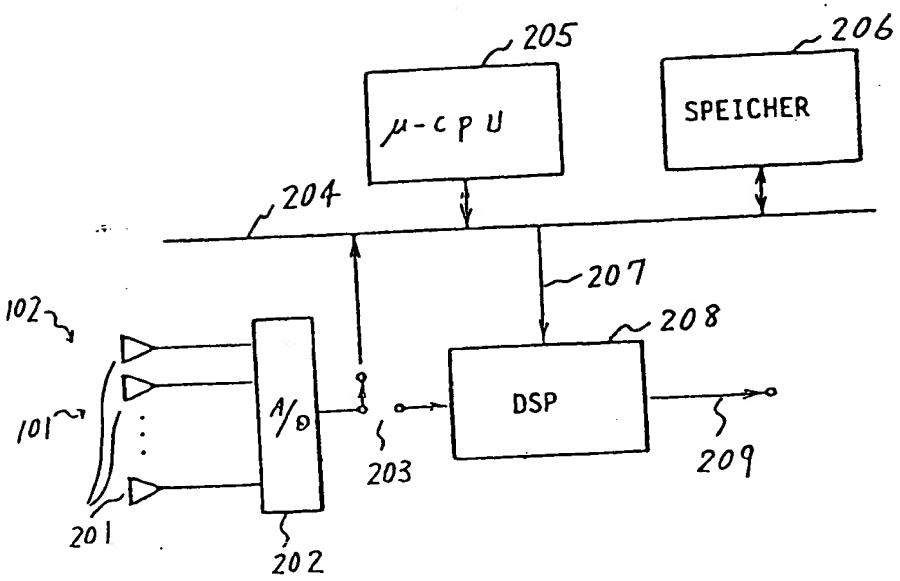
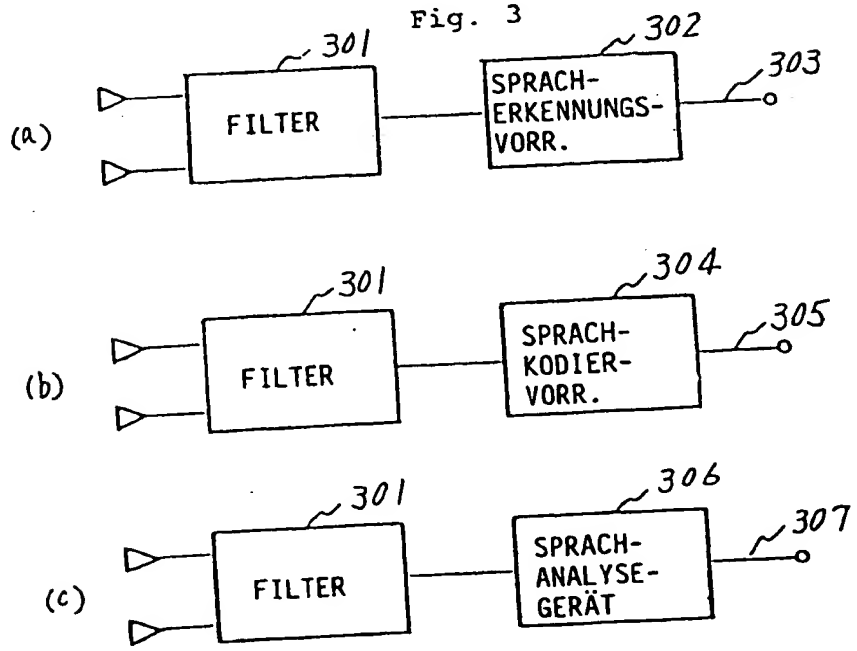


Fig. 3



Noise reduction equipment in speech signal processing system - couples microphones by hierarchical neuronal networks to digital processor

Patent Number : DE3929481

International patents classification : G10L-003/00 G10L-005/00

• **Abstract :**

DE3929481 A A number of spatially separate transducers produces a number of electric speech signals corresp. the sound acting on them. A reusonal network has a number of layers each comprising a number of reusonal elements, and there being an input layer and an output layer.

A first communication device transmits the electric speech signal from a first of the transducers to each element of a first of reuson elements in the input layer. A second communication device transmits the electric speech signal from a second of the transducers to each element of a second set of reuson elements. Each communication device may convert the analog speech input to digital and may include a shift register for providing separated outputs.

ADVANTAGE - Improved speech clarity and comprehension. (o.3/3)

DE3929481 C The device described is based on a number of converters (103,104,105), arranged some distance apart, producing a number of electrical speech signals, corresponding the sounds impinging on them. The significant feature is a neuron-like network with a number of layers (114,117,119,121), including an input (114) and an output (121) layer and each layer consisting of a number of neuron elements (115,116;118; 120;122). These receive the speech and noise signals and, through a monitored learning process, weighting data is obtained, establishing the transmission properties of the network. There are two transmission units (106,112;107,113), a first and a second, for transmitting the electrical speech signals to a first and a second set of neuron elements in the input layer. USE/ADVANTAGE - Improvement in suppression of background noise and so in recognition of speech. Advanced computer and robot techniques. 4/7/91 (8pp)

• **Publication data :**

Patent Family : DE3929481 A 19900315 DW1990-12 8p * AP: 1989DE-3929481 19890905
DE3929481 C 19910704 DW1991-27
Priority N° : 1988JP-0222309 19880907
Covered countries : 1
Publications count : 2

• **Patentee & Inventor(s) :**

Patent assignee : (HITA) HITACHI LTD
Inventor(s) : AMANO A; ASAKAWA Y; HATAOKA N; ICHIKAWA A

• **Accession codes :**

Accession N° : 1990-084690 [12]
Sec. Acc. n° non-CPI : N1990-065367

• **Derwent codes :**

Manual code : EPI: T02-A04A9 W04-G W04-V
Derwent Classes : P86 T02 W04

• **Update codes :**

Basic update code :1990-12
Equiv. update code :1991-27